A peer-reviewed open-access journal
NeoBiota
Advancing research on alien species and biological invasions

NeoBiota 105: 177–208 (2026)
DOI: 10.3897/neobiota.105.159101

Research Article

Streamlining alien species listing processes to enable prioritisation and reporting in protected areas: The case of alien plants in South African National Parks

Nicola J. van Wilgen[1,2], Emma J. Wright[1], Johan Baard[3], Bevin Ahrends[4], Nicholas Cole[4], Chad Cheney[5], Llewellyn C. Foxcroft[2,6], Dian Spear[1,7]

1 *SANParks Cape Research Centre, Tokai Road, Tokai, 7945, South Africa*
2 *Centre for Invasion Biology, School for Climate Studies, Stellenbosch University, Stellenbosch, South Africa*
3 *SANParks Garden Route Scientific Services, Knysna, 6570, South Africa*
4 *SANParks Biodiversity Social Projects, Cape Town, South Africa*
5 *SANParks Table Mountain National Park, Tokai Road, Cape Town, South Africa*
6 *SANParks Scientific Services, Skukuza 1350, South Africa*
7 *Centre for Sustainability Transitions, Stellenbosch University, Stellenbosch, South Africa*
Corresponding author: Nicola J. van Wilgen (nvanwilgen@gmail.com)

OPEN ACCESS

This article is part of:
Developing lists of alien taxa in the Global South: workflows, protocols, processes, and experiences
Edited by John Wilson, Michele Dechoum, Katelyn Faulkner, Barbara Langdon, Shyama Pagad, Aníbal Pauchard, Hanno Seebens, Tsungai Zengeya, Silvia Ziller

Academic editor: Tsungai Zengeya
Received: 17 May 2025
Accepted: 19 December 2025
Published: 22 January 2026

Citation: van Wilgen NJ, Wright EJ, Baard J, Ahrends B, Cole N, Cheney C, Foxcroft LC, Spear D (2026) Streamlining alien species listing processes to enable prioritisation and reporting in protected areas: The case of alien plants in South African National Parks. NeoBiota 105: 177–208. https://doi.org/10.3897/neobiota.105.159101

Copyright: © Nicola J. van Wilgen et al. This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Abstract

Management of alien species is a critical conservation priority in protected areas. Reliable alien species checklists are central to prioritisation, early detection, long-term monitoring and reporting. Yet, methods vary widely, from structured field surveys with herbarium vouchers to data aggregated from online platforms, atlases, citizen science and scientific journal publications. These sources often use inconsistent naming systems, especially where alien species names differ between the introduced and native range. We describe the processes used over three decades by South African National Parks to compile alien species lists, highlight key shortcomings and propose a ten-step process to improve checklist development. Our approach generates a dynamic occurrence database that integrates static historical datasets with dynamic sources, such as iNaturalist, from which up-to-date checklists can be produced. Central to this process is taxonomic standardisation against a taxonomic backbone, combined with annotations of identification certainty and spatial accuracy and the assignment of unique record and taxon identifiers. Applying this process across parks yielded a consolidated list of 1,330 unique taxa – significantly longer than previous versions. The increase reflects broader scope (e.g. inclusion of 206 extralimital taxa) and incorporation of new data streams, while some park lists shortened due to removal of misidentified or synonymous taxa. Around one-third of alien species recorded were absent from the national checklist, underscoring the need for a systematic approach to alien species listing. As conservation indicators increasingly depend on such lists, we recommend explicit guidance is provided on: (1) classifying invasion status and stage, especially for cultivated species; (2) treatment of extralimital aliens; and (3) verification standards for non-expert or spatially imprecise records. Using our proposed process alongside such guidelines will provide a robust foundation for consistent alien plant species reporting in protected areas, strengthening the value and application of alien species lists.

Key words: Conservation targets, GBIF, iNaturalist, invasive species, taxonomic backbone, taxonomy

Introduction

With the acceleration of threats to natural systems by global change, areas of high biodiversity and endemicity have been flagged as high risk for biodiversity loss (Manes et al. 2021). Alien plant and animal species are a key threat that pose diverse and significant negative impacts for people and the environment globally (van Wilgen et al. 2022a; IPBES 2023). The Convention on Biological Diversity's Kunming-Montreal Global Biodiversity Framework has established clear targets (Targets 2, 3 and 6) aimed at increased area under conservation, restoration of degraded ecosystems and preventing the introduction and spread of alien species, emphasising the need for effective management strategies within protected areas (CBD 2022; Watson et al. 2023). South Africa presents a useful case study, as it is unique not only in its exceptionally high biodiversity, which includes three of the world's biodiversity hotspots (Myers et al. 2000), but it is also one of the top five most diverse countries for alien species globally (Turbelin et al. 2017). The changing climate, diminishing management budget and habitat fragmentation are increasing pressure on the management of natural systems and complicate existing trade-offs in species and area prioritisation (van Wilgen et al. 2020; Evans et al. 2025). Managing these trade-offs effectively requires data on the presence, distribution, abundance, spread and impact of alien species.

Protected area agencies need to prioritise alien species management both within and between parks (Cheney et al. 2020a). Considerations include the prevention of impacts in high biodiversity areas or areas where other assets are threatened (e.g. infrastructure or tourism value), cost, ease of access and availability of effective control measures (Roura-Pascual et al. 2009). As such, the maintenance of comprehensive checklists containing essential biodiversity data is imperative (Reyserhove et al. 2020). Such checklists form a critical baseline for comparisons and facilitate informed decision-making regarding resource allocation for conservation, species monitoring and targeted management of rare or invasive species (Droege et al. 1998; Rebelo et al. 2011; Hamer et al. 2012). However, inconsistency, along with incomplete, inaccessible or unverifiable data (e.g. absence of necessary metadata) and the lack of taxonomic expertise (Pyšek et al. 2013), severely constrain the completeness and usefulness of potential data sources (Stephenson et al. 2017). In addition, despite available data sharing standards like Darwin Core and data platforms, such as the Global Biodiversity Information Facility (GBIF), which is underused in checklist creation (Spear et al. 2026), many conservation organisations lack the capacity to curate species lists due to knowledge, technology, resource and data-sharing barriers (Wieczorek et al. 2012; Gaiji et al. 2013).

There is a need for a simple, accessible and cost-effective approach to improve data accessibility and utility within protected areas. One of the greatest stumbling blocks for managing species data is an inability to match species names across areas and over time (Spear et al. 2023; Fig. 1). For alien species, this is further complicated by misapplied names, as the common and scientific names used for species in their country of origin differ from those used in the areas where they have been introduced (Fig. 1). Taxonomic backbones are essential for resolving name-matching issues (Thomson et al. 2021), but several limitations remain that require careful choice and transparent reporting to develop dynamic checklists. Firstly, different backbones may be required for native versus alien species.

For example, in South Africa, the South African National Plant Checklist (SANPC, SANBI 2025) provides a curated list for native plants, but does not comprehensively include aliens. This phenomenon is widespread globally, as curation of species lists for native species is often approached with a systematic ecological or conservation focus (e.g. Carrero et al. 2022), whereas alien species lists tend to be legislative and are updated as problem species arise or include only certain classes of alien species, such as those that are invasive (Stohlgren et al. 2011; Danihelka et al. 2012; Natural Resources Conservation Service 2025). To address this, Spear et al. (2023) recommended using the GBIF taxonomic backbone, which largely follows the World Checklist of Vascular Plants and includes all species that would be native and alien to the region of interest. Secondly, a significant consideration is the extent to which name matching can be automated. Some authors have combined automated tools (e.g. the Taxonomic Name Resolution Service) with manual alignment (e.g. Wenk et al. 2024). A key advantage of the GBIF backbone is that outdated names can be uploaded and matched and it offers a single reference system that can be applied across taxonomic groups, an important feature for protected area databases (Spear et al. 2023). However, for specific taxa, there are several other considerations during taxonomic harmonisation, which may affect the choice of backbone and related tools (Grenié et al. 2022).

As with the compilation of any species checklist, compiling alien species checklists inevitably involves collating and validating data and standardising names (Spear et al. 2023). Data need to be collated from scientific publications, herbarium records, field surveys and data platforms, such as the Global Biodiversity Information Facility (Reyserhove et al. 2020) or Plants of the World Online (Govaerts 2025) [see also Sankaran et al. (2023) for an overview of resources]. Data also need to be checked for identification and location accuracy (Spear et al. 2023). However, static checklists rapidly become outdated due to shifts in taxonomy and the changing distribution of alien species, limiting their long-term value and integration with new data. While numerous protocols for generating checklists exist, there are very few worked examples that are broad in scope, practical in application and attentive to challenges across multiple phases of checklist creation. This gap is critical, as protected areas urgently need tools that can standardise data across diverse sources to support reporting, monitoring and management. In this paper, we present a systematic and replicable methodology for integrating heterogeneous data sources (e.g. historic collections, field research and contemporary online platforms), into dynamic species lists that can be applied in protected area species listing and prioritisation processes. Our approach is unique in explicitly addressing both the technical and practical challenges of checklist development within the context of biological invasions in protected areas, highlighting the importance of iterative updating and refinement as new information emerges (Fig. 1). By doing so, we provide not only a comprehensive worked example, but also a scalable resource to guide prioritisation, early detection, long-term monitoring and reporting. This approach not only represents a major step forward for South African National Parks (SANParks), who operate across eight biomes and engage with a wide national network of data generators and users, but it can also be adapted to other areas that span diverse contexts, such as the European Union or the United States of America. Our case study focuses on alien plants, but the approach is generally applicable.

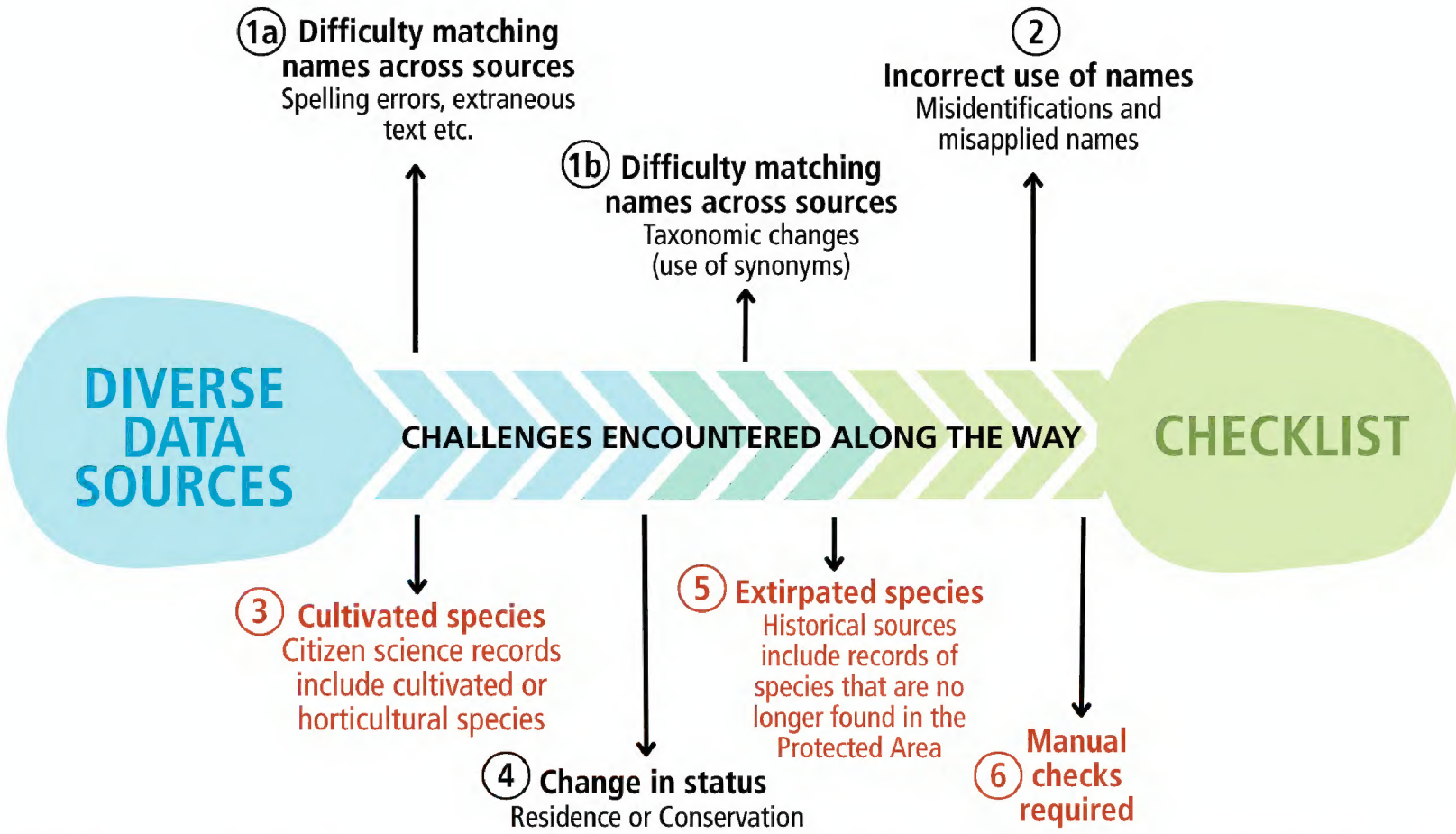
1a Difficulty matching names across sources
Spelling errors, extraneous text etc.
2
Incorrect use of names
Misidentifications and misapplied names
1b Difficulty matching names across sources
Taxonomic changes (use of synonyms)
DIVERSE DATA SOURCES
CHALLENGES ENCOUNTERED ALONG THE WAY
CHECKLIST
3 Cultivated species
Citizen science records include cultivated or horticultural species
4 Change in status
Residence or Conservation
5 Extirpated species
Historical sources include records of species that are no longer found in the Protected Area
6 Manual checks required

Figure 1. Challenges encountered when compiling species lists from diverse data sources. Our paper provides a process that addresses the challenges in black, while the challenges in orange require a more nuanced approach and are covered in more detail in the Discussion.

Materials and methods

South African National Parks is fortunate, compared to many protected areas globally, that the organisation has many in-house researchers (Roux et al. 2019). While research effort by these researchers is not evenly distributed, with most research occurring in the parks with better research infrastructure, there are several parks where dedicated botanical studies, including studies on alien plant diversity, have been undertaken (Foxcroft and Richardson 2003; Foxcroft et al. 2008; Masubelele et al. 2009; Baard and Kraaij 2014; Cheney et al. 2018; Foxcroft et al. 2023). Two comprehensive species lists have been compiled across parks (Spear et al. 2011; Foxcroft et al. 2017), based on all available information at the time, using a variety of source types, but these lists are not easily updatable due to several challenges (Fig. 1). Firstly, the names used were not complete names inclusive of authors, subspecies and varieties (where relevant). Names were also not unambiguously associated with sources or unique taxon identifiers from a taxonomic backbone. As such, names could not be matched to a particular source in all instances, nor could they be matched across sources. It was not clear whether new names from diverse sources represented new taxa, misidentifications or merely name changes. We have used lists from the previous studies as a basis for understanding how species checklists were compiled in the past, while also serving as an important source of information for trialling our proposed process. We followed a 10-step process to enable the generation of a dynamic database from which a living species checklist (i.e. one that can be easily reproduced and updated) can be drawn. The same process is applicable to the curation of data for native species and can be refined through the definition of the scope in Step 1 (Fig. 2).

Defining scope (Step 1)

The first step of the process is to define the geographic and taxonomic scope of the list. In this instance, we were interested in listing all alien plant species across national parks in a country, i.e. 21 national parks in South Africa

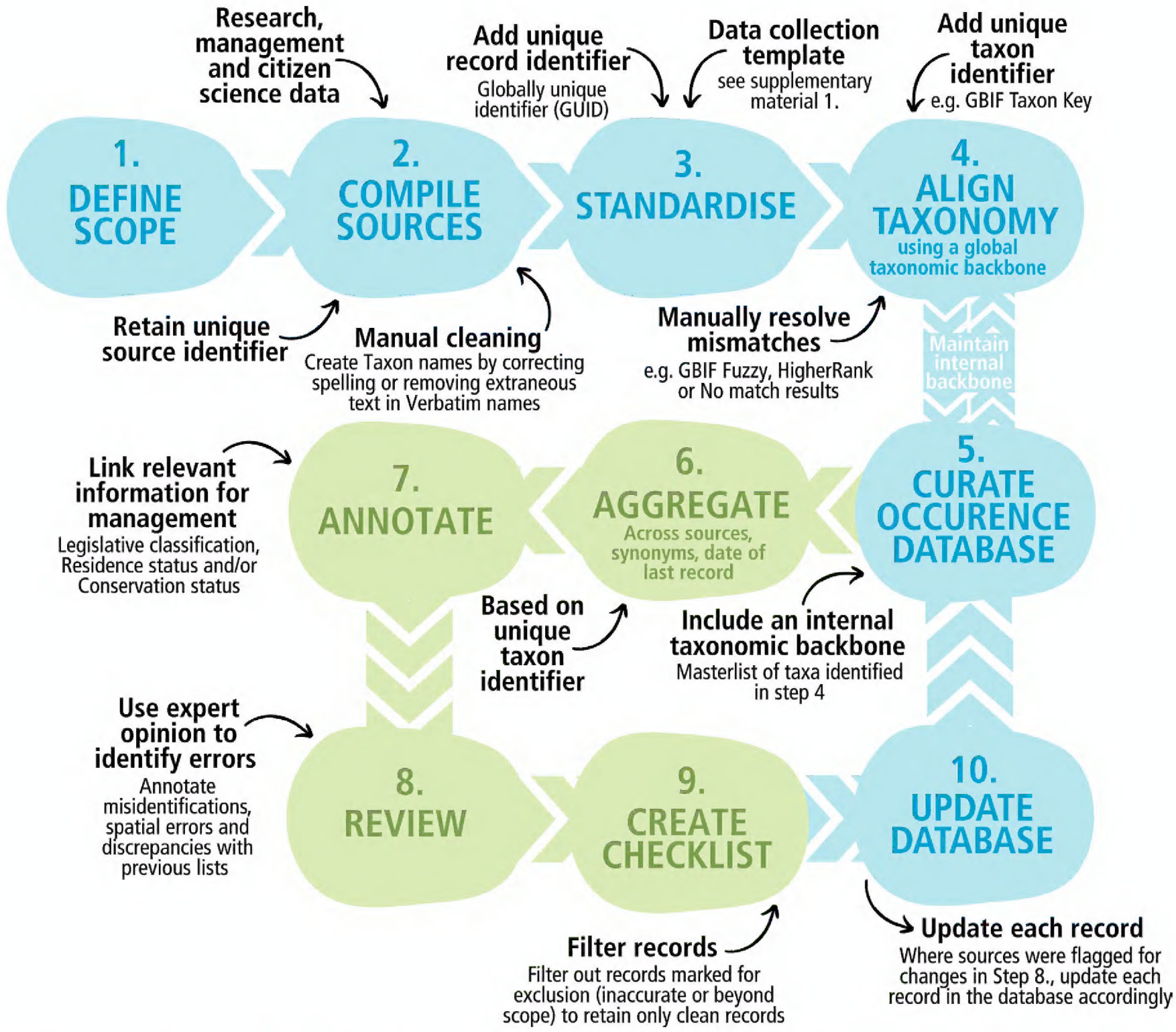
Research, management and citizen science data
Add unique record identifier
Globally unique identifier (GUID)
Data collection template
see supplementary material 1.
Add unique taxon identifier
e.g. GBIF Taxon Key
1. DEFINE SCOPE
2. COMPILE SOURCES
3. STANDARDISE
4. ALIGN TAXONOMY
using a global taxonomic backbone
Retain unique source identifier
Manual cleaning
Create Taxon names by correcting spelling or removing extraneous text in Verbatim names
Manually resolve mismatches
e.g. GBIF Fuzzy, HigherRank or No match results
Maintain internal backbone
Link relevant information for management
Legislative classification, Residence status and/or Conservation status
7. ANNOTATE
6. AGGREGATE
Across sources, synonyms, date of last record
5. CURATE OCCURENCE DATABASE
Based on unique taxon identifier
Include an internal taxonomic backbone
Masterlist of taxa identified in step 4
Use expert opinion to identify errors
Annotate misidentifications, spatial errors and discrepancies with previous lists
8. REVIEW
9. CREATE CHECKLIST
10. UPDATE DATABASE
Filter records
Filter out records marked for exclusion (inaccurate or beyond scope) to retain only clean records
Update each record
Where sources were flagged for changes in Step 8., update each record in the database accordingly

Figure 2. The proposed process for collating alien species lists from diverse data sources. Steps in blue relate to developing and maintaining an occurrence database, while steps in green describe the creation of a checklist from the database. Depending on the (1) scope and (2) source of data, datasets are compiled and Verbatim names cleaned into recognisable Taxon names (e.g. correcting spelling or removing extraneous text). Each source is standardised using a set data collection template (3) that provides a unique record identifier for each occurrence record. All names are then standardised to align taxonomy using a global taxonomic backbone (4), with mismatches resolved to provide a unique taxon identifier for the accepted version of names. Standardised, cleaned and matched datasets are then curated into an occurrence database (5). A preliminary species list can be drawn from this database by aggregating taxa by the unique taxon identifier across sources and synonyms (6). The list can be annotated with management relevant information by matching to other tables, based on taxon identifiers (7) and expert review sought to identify errors and inaccuracies (8). An up-to-date, user-specific checklist can be generated (9) by filtering out records marked for exclusion during any of the previous steps. In the absence of a dynamic occurrence database, excluded records should be retained with appropriate annotations or maintained in a parallel database. Ultimately, all changes should be annotated per record, which requires flagging sources where changes are required and updating these within the occurrence database (10) and where relevant at source (e.g. where the source is digital).

(Fig. 3) and additionally considered species present in the water catchment areas around parks for use in watchlists (see Discussion). Given the diversity of biomes in South Africa, we considered both species alien to the country and species extralimital to each park, where this information was available. Extralimital species are those that are native to parts of South Africa, but have been introduced to new areas (Nelufule et al. 2022). These species interact in new areas in the same way as other alien species, while also having the potential to hybridise with closely-related species.

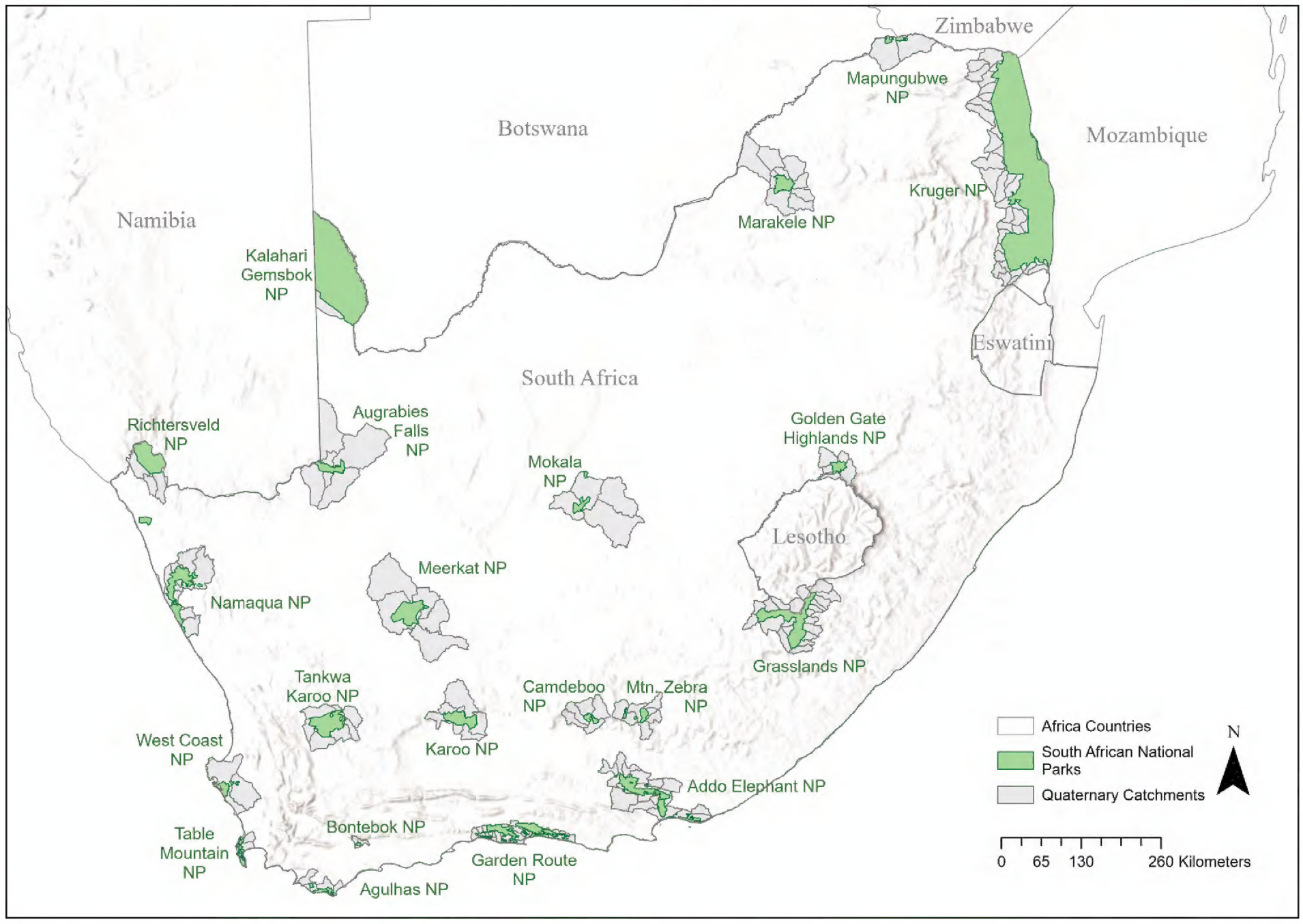

Figure 3. Location of the 21 national parks managed by South African National Parks (green shading) and the area around them connected to quaternary catchments (grey shading) from which additional iNaturalist data were collected. These catchments represent conservative buffer zones, but could be amended as desirable according to relevant pathways.

Compiling data from diverse sources (Step 2)

In Step 2, relevant data sources are identified. Typically, occurrence data have been collected for one of three primary purposes: for describing a flora, i.e. for identification and listing of species presence; for management; or for research. An emerging source is citizen-science data. Within these categories, the level of taxonomic expertise of data collectors varies, for example, little taxonomic expertise is required for citizen-science data collection. Not all data are readily usable for checklists and may require different levels of effort to ensure their usability. While using existing lists as a source is often the most convenient option, obtaining information directly from original sources where possible is advisable, as this provides context for the names used and the manner and expertise involved in collecting the data. Where source data contain unique identifiers, it is important to retain these for each record to enable dynamic links between source data and your occurrence database. For example, a research-grade record from iNaturalist may change as contributors add identifications. If the unique identifier associated with that record on iNaturalist is not retained, then any identification updates would be captured in your occurrence database as new records, rather than updating the original record with new confidence information. Here, we describe the data available in these categories within SANParks.

Data collected by botanists specifically for describing flora

Field surveys are the primary method for verifying species presence and abundance across habitats and capturing data on newly-emerging invasions. For example, in 1998, in-house scientists in what is now the Garden Route National Park took the decision to prepare a species list of all plants for parks in forest habitat. Existing publications (primarily Geldenhuys 1993) were used as the basis of compiling the list. However, the list was not comprehensive and no alien species or fynbos species were recorded. A herbarium was established in 1998 at the Knysna Department of Water Affairs and Forestry as a reference for identifying new species (Smith and Willis 1999). An alien species list was then compiled, starting with field observations of species for which the taxonomy was clear or easy to determine. These were often widely known invaders and observations of species kept as ornamental plants, many of which 'escaped' into surrounding natural areas (e.g. *Quercus* species, *Cinnamomum camphora*). An important factor for identifying alien plants was that staff had developed a good institutional memory of the naturally occurring species of the region. When unfamiliar species were suspected of being alien, they were sent to local herbaria (e.g. Compton Herbarium, https://www.sanbi.org/biodiversity/foundations/biosystematics-collections/compton-herbarium/) for identification. In this way, even species that had been assumed to be native were identified as alien. An important aspect of this approach was the proper collection of specimens for taxonomic identification, as described in Victor et al. (2004).

A herbarium was established by biologists in Skukuza, Kruger National Park, around 1950, with the aim of creating a plant inventory for the park. This herbarium now includes approximately 2,000 species and more than 15,000 specimens contributing to plant identification and providing a reference of the plant species of the park (Singh et al. 2025). Alien plant species were included in the collection from the outset. Other national parks have species inventories and voucher specimens housed at the Kimberly herbarium, with the same aims, amongst others, of assisting with the identification of alien plant species (Singh et al. 2025).

At a national level, alien species 'field guides' and specialised web sites were developed to assist in the identification of alien species (e.g. Henderson 1995, 2001; Henderson and Cilliers 2002; Bromilow 2010; Walters et al. 2011; Bromilow 2014; Henderson 2020; Klein and Zimmermann 2020). However, even with field guides, there are some species and species groups that may still require expert assistance for identification (e.g. *Callistemon*, *Melaleuca* and various *Myoporum* species). With the advent of platforms like iNaturalist, opportunities for receiving expert feedback rapidly is feasible. Botanists can upload photographs that match required identification keys (Blakely 1955; Victor et al. 2004) and ask their peers for online feedback on the platform, a practice that is growing. Online databases, such as Plants of the World Online (POWO) and the Global Invasive Species Database (Pagad et al. 2015), may also assist in narrowing identification, as they contain photographs and morphological descriptions.

Field surveys by non-experts

In some instances, it may be necessary to conduct surveys without the expertise of botanists or other experts. Some taxa, like Cactaceae, are easy to identify as alien, as there are few similar species in the natural environment (although *Euphorbia*

species are occasionally incorrectly identified as cacti, Anderson 2001). Other taxa, like grasses and Cyperaceae, are often easily overlooked, as well as being difficult to identify. In protected areas, park rangers may be responsible for managing alien species within pre-defined management zones, but may not be familiar with all the species in these areas. In most instances in South Africa, clearing programmes make use of unskilled or semi-skilled labour. People involved in clearing are typically taught to identify a few key species (and species to avoid clearing), but would not be able to distinguish native from alien species in general. Similar situations may exist where protected areas make use of volunteers in clearing programmes. Field guides, particularly those tailored to specific regions or groups (e.g. Walters et al. 2011) can be very useful for training non-experts in identifying alien species in the field, in particular to identify the correct family or genus, from where more detailed investigations can be done. Guides that describe commonly confused species (e.g. Bromilow 2014; Henderson 2020) and include line drawings specifying character-istics to look out for are particularly helpful (e.g. Henderson 2001). For example, in Table Mountain National Park, a large-scale survey of all alien plants was conducted across 10,000 survey plots by non-experts (Cheney et al. 2018), who gained experience in identifying plants and were able to produce a valuable dataset with high identification confidence (Cheney et al. 2020b; Cheney et al. 2021).

Data platforms, like iNaturalist, have significantly expanded the potential for non-experts to make contributions. The identifications on such platforms are, however, only as reliable as the photographs and descriptions provided by the users of the platform and the users doing the identification. Selecting research-grade identifications improves the quality of the data significantly (Wenk et al. 2024). In general, however, data from non-experts require additional cleaning and verification compared with formal surveys conducted by botanists and verified by taxonomists. For example, even research-grade iNaturalist data only ensure that two users have agreed to an identification, while providing little insight into the training and skill of the identifiers (Ackland et al. 2024).

National management data

Much of the data on alien plants that has been used in reporting on the impact of alien clearing in South Africa (e.g. van Wilgen et al. 2022b) have been derived from the Working for Water Information Management System (WIMS), a database that has been used to track and manage all the government-funded alien clearing contracts in South Africa since 2003. The programme does not provide an exhaustive record of species, as it only focuses on a few key species that are treated and does not document the less common alien species (Cheney et al. 2018). During the early stages of adopting the system, many species were incorrectly captured, due to spelling errors and misidentification in field, resulting in inaccuracies and duplication. Additionally, species were grouped by genus or clearing requirements, meaning that sometimes contracts for clearing work were generated, based on requirements for commonly cleared species, such as *Acacia* or *Pinus* species, but the species being cleared may have been a *Hakea*, *Grevillea* or a *Banksia* with similar clearing protocols, potentially resulting in errors when generating location-based species lists from this system. The static nature of historical names in WIMS complicates analyses incorporating these data to assess trends, as there is no clear method to match species names that have been added or corrected over time

(e.g. van Wilgen et al. 2022b). One of our aims was to review historical inaccuracies in this system and, where possible, annotate corrections and updates to enable better cross-reference of programme outcomes over time and across parks.

To illustrate Step 2, we took the existing list compilations (Spear et al. 2011; Foxcroft et al. 2017) and identified the original data sources as a basis of alien plant species records. From these lists and our recent work, we curated data from the following sources: Foxcroft et al. (2003); Foxcroft et al. (2008); Masubelele et al. (2009); Baard and Kraaij (2014); Baard and Kraaij (2019); as well as data from separate unpublished databases of species curated by Johan Baard for Garden Route National Park and Chad Cheney for Bontebok National Park; Cheney et al. (2020b); Keet et al. (2022); Management data from WIMS (SANParks 2024), as well as other individual sources cited within Spear et al. (2011).

Citizen-science data

Citizen science is able to gather large amounts of data inexpensively and is a key source for detecting emerging invasions (Larson et al. 2020), though challenges remain with data curation and reliability for use in management (Wenk et al. 2024). We sought to determine how the inclusion of iNaturalist data might complement our park species lists. We received iNaturalist data from the South African National Biodiversity Institute (SANBI - the national biodiversity data curator) that included all records of alien species in parks and their adjacent quaternary catchment areas (Fig. 3) up until the 14 July 2023 (n = 181,395). We then drew a further 25,170 research-grade records from locations within parks and their adjacent quaternary catchment areas for the period 15 July 2023 to 18 March 2024. Only research-grade records were included in our final list.

For all datasets, the name by which a taxon was referred to in the source was stored as is under "Verbatim name". However, sometimes names included long descriptions or obvious spelling errors. Where possible, these were refined into recognisable "Taxon names" by removing extraneous text and correcting spelling.

Standardising and aligning data (Steps 3 & 4)

Step 3 involves setting up a data collection template to standardise data and metadata for each source dataset and enable consistent data capture with the purpose of later compiling an occurrence database (Fig. 2, Step 3). For our purposes, we based this template on the requirements of the SANParks Biodiversity Information Management System (BIMS) (https://bims.sanparks.org/), a dynamic spatially enabled occurrence database (see Suppl. material 1: Data collection template). The BIMS was developed as a platform for storing species occurrence data from national parks. In preparing the source data, each record was given a globally unique record identifier using a built-in algorithm (Fig. 2, Step 3). Any unique identifiers from the record's source as well as full details of the source were captured to ensure the ability to cross-reference into the future. Other fields of importance include species' residency, including invasion status at that location (Richardson et al. 2000) and information on the accuracy of the record (see suppl. material in Spear et al. 2023 for recommended fields to use). Different sources inevitably have diverse use of taxonomy and adherence to principles of plant nomenclature, which necessitates a means of aligning taxonomy.

Step 4 therefore involves identification of an appropriate global taxonomic backbone to align taxonomy across records and sources, by providing a unique taxon identifier for the accepted name of Taxon names. The resources available for taxonomic clarification have diversified over the past decades; for example, the use of names can be checked on platforms like Catalogue of Life or Taxonomic Name Resolution Service (Grenié et al. 2022). The diversification of these resources, however, means that discrepancies in accepted names occur. In these instances, using a nationally accepted name is preferable, but in the case of alien species, national taxonomic backbones do not necessarily include all alien taxa or may use a locally understood name that has been changed in the species' native distribution. For example, a species like *Agrostis avenacea* C.C.Gmel. has been assigned to the genus *Lachnagrostis* in its native range (*Lachnagrostis filiformis* (G.Forst.) Trin.), but *Agrostis avenacea* is the name on the national backbone (SANPC). While Zengeya et al. (in press) propose an alternative way for handling names that are not included in the national checklist, using alternative backbones only where necessary, we propose using a single global backbone as a more streamlined solution. As BIMS uses GBIF as a backbone, which is based on POWO, the source of the World Checklist of Vascular Plants, also used by iNaturalist, this shared primary reference was an obvious choice for aligning source data taxonomy. In addition, the GBIF backbone can also be used for animals, therefore requiring only one system for protected area occurrence data across taxa. We used GBIF's species lookup tool in the R package *rgbif* (Chamberlain et al. 2024) to identify accepted names for all the names used across our sources. The name matching function (name_backbone) provides an accepted name associated with the Taxon name (name cleaned from the source), along with a unique taxon identifier known as a taxonKey (Fig. 2, Step 4) and indicates the taxonomic level at which the match has been made (e.g. subspecies, species, genus).

While each name matching tool will present its own challenges, using the matching tool available for the GBIF backbone provided only the scientific name for the Taxon name provided, along with its status (accepted, synonym etc.) and the taxonKey for this name. Where the matched name was a synonym, it did not provide the taxonKey for the Accepted name. We modified our data queries in R (R Development Core Team 2024) to provide the accepted name where the matched name was a synonym, but this required a second query of the GBIF database. The latter could be avoided if the lookup function was updated to produce both the scientific name of the synonym and its associated accepted name along with the taxonKeys for both these names. Our application of the *rgbif* function was also not able to implement comprehensive fuzzy matching and, in many instances, provided the accepted name for a higher taxonomic rank than the Taxon name provided. Due to our large dataset and intensive computing requirements, standard queries took a long time (hours to days) to process. Therefore, we modified our code with the assistance of ChatGPT (OpenAI 2025). This allowed for parallel processing while also maintaining a running list of names that have already been matched, preventing re-running of names that had already been checked, reducing the run time to minutes. See Suppl. material 2: R code for checking species names, for functions compiled in R (R Development Core Team 2024) to collect accepted names and taxonKeys.

Following matching, further manual data cleaning is required. Despite some automated data cleaning, there were some names that were detected by the matching

tool that required manual searching for an accepted name and taxonKey. The GBIF output provides a "Match type". All instances where the match type for the matched Taxon name was classed as Fuzzy, HigherRank or NA were manually assessed. Where matches were incorrect, no match was made (NA) or a GBIF HigherRank match was made, manual searches on GBIF were required to add the correct accepted name and taxonKey. Where a manually matched name was a synonym, the appropriate accepted name and taxonKey were also recorded.

Curation into an occurrence database (Step 5)

In Step 5, data are collated across sources into a relational occurrence database. Depending on available resources, an occurrence database could range from a simple Excel spreadsheet linked across tabs using lookup functions to an internal information management system (e.g. SANParks' BIMS) or an open-data platform, such as GBIF (see Reyserhove et al. (2020), for using GBIF to curate checklists). However, the most basic components of such a relational occurrence database must always include an internal taxonomic backbone of all taxa that exist within the database and the occurrence data associated with each taxon from all data sources. The internal taxonomic backbone should be kept up to date with the addition of any new taxa. The taxonomy in the database should be updated at regular intervals, ensuring that the accepted names for species remain current and assessed by matching the unique taxonomic identifier between the internal taxonomic backbone and the reference global backbone (e.g. here the GBIF backbone). In the absence of a formal database, maintaining files of cleaned source data with relevant identifiers is crucial. Erroneous records should not be removed, but rather flagged as such in the appropriate column (taxonomic, spatial or identification accuracy). For our purposes, all cleaned and annotated, unsummarised locality data will be uploaded to SANParks' BIMS. Within this system, internal taxonomic backbones will be updated to include all Taxon names used across sources. Taxonomy within the internal backbones will be maintained current using regular alignment to the latest GBIF taxonomic classification.

Aggregating, annotating and reviewing the checklist (Steps 6–8)

In Step 6, a preliminary checklist can be created by aggregating data to retain one record for each species, collating all the sources that listed the species together with additional information of interest recorded in the source data (e.g. date last sighted, synonyms). We did this using dplyr (Wickham et al. 2023), summarising functionality to group taxa, based on taxonKey and aggregating a list of the sources and synonyms for each key.

In Step 7, the list can be annotated with user-defined information using linked tables. For example, it may be useful to include legislative categories, species status information or other management information relevant at a species level, such as key pathways. If applying a similar approach for native species, one might want to add information on species' conservation status. In South Africa, regulations under the National Environmental Management: Biodiversity Act (no. 10 of 2004, hereafter NEMBA) provide a list of alien species for which control is mandatory, within three categories. We matched the species listed on NEMBA to GBIF to provide accepted names and taxonKeys for each species (Steps 4) and joined this table with our species

list table in R by matching taxonKeys (see Suppl. material 2, in conjunction with Suppl. material 3: NEMBA list categories). Through this process, fine levels of detail can be added. For example, certain species listed have different control requirements in different administrative provinces. Appropriate provincial control requirements were matched according to the administrative location of each park.

A limitation of any name-matching tool is that misidentifications and spatial errors will not be identified. Step 8 entails reviewing the preliminary annotated list generated in Step 7. We manually reviewed the consolidated list for discrepancies with previous lists (Spear et al. 2011; Foxcroft et al. 2017). Where names had been marked as incorrect in the previous revision, records were flagged for removal with reasons as appropriate. The new annotated lists (overall and per park) indicating all species that were not on previous park lists, as well as those marked for removal from lists, were circulated to experts, where available, to comment on accuracy and identify potential misidentifications or concerns around the use of names (Step 8). We focused our verification checks on WIMS data because several known problems were associated with this source. The process of thoroughly checking each data source for misidentifications was not possible within the scope of the current project, where our focus was to trial the method. However, listing the source information for each species record allows users of the lists to make their own inferences about data accuracy and flag records for exclusion where obvious errors are identified. In addition, our cleaned raw data are associated with taxonKeys in the occurrence database that will enable more dynamic use of the data in future list updates.

Production of a checklist (Step 9)

A final checklist is produced by filtering out records marked for exclusion during any of the previous steps. In the absence of a dynamic occurrence database, a minimum requirement would be the retention of incorrect records, appropriately annotated or maintained in a parallel database. The parallel database of incorrect species records is required to prevent inadvertently re-including records that had been evaluated as inaccurate. Depending on how protected areas plan to maintain their data, Step 9 may be a once-off static checklist.

Updating the database (Step 10)

To ensure a living checklist, the final step (Step 10) ensures that all changes in identification, certainty or related attributes are annotated directly within a relational occurrence database or dynamic system such as BIMS. As Step 8 is based on aggregated data, sources where errors occur need to be flagged and all relevant occurrence records updated. To provide the required functionality, the system must accommodate updates to species names and expert-suggested changes, sometimes long after the original data were collected. For example, retrospective action is required where names have been misapplied. To preserve traceability, the original Verbatim name from the source is retained, while the Taxon name is updated to reflect current expert understanding of what the record represents. The accepted name for this Taxon name can be kept up-to-date through regular matching to a taxonomic backbone. Confidence estimates in identifications (identification accuracy) are updated in parallel to track updates in the Taxon name. This structure allows for the generation of customisable, up-to-date checklists that are meaningful

for tracking invasion patterns over time. Ultimately, the way records are linked and updated across the source, occurrence database and taxonomic backbone determines the long-term utility of any resultant lists (Fig. 2).

For example, changes or annotations to national management data sources are not always feasible. In this case study, drawing data from WIMS re-introduced errors that had been amended in previous lists. To address this, we created unique identifiers for each WIMS record across park–species–site–date to enable tracking of changes between the static WIMS species list and the SANParks BIMS. The exported WIMS data were cleaned and annotated to flag name corrections and misidentifications, producing a dataset that accounts for historical inaccuracies up to the end of 2024. This dataset will be uploaded into the SANParks BIMS, while an intermediary file has been retained to link historical data between databases, along with a workflow for applying comparable corrections to future data (see Suppl. material 4: Cleaned WIMS data and workflow). Where source data are also drawn from a dynamic system (e.g. GBIF or iNaturalist), additional data integration challenges may arise as multiple systems are implementing changes to data using different platforms and software. When records are updated in the source data, a mechanism is required to reflect these updates in the occurrence database to ensure that inaccurate legacy data do not remain in the system. While not all systems will be able to automate this (it may be a manual task), it is something that database managers should be aware of.

Data analysis

Summary statistics per park were compiled using summarising functionality in R version 4.4.2 (R Development Core Team2024) and Excel. To assess the utility of the national backbone for alien plant species listing, we matched the names used in the national backbone (SANBI 2025) to GBIF to identify taxonKeys for the accepted version of names and then matched these to the accepted names from our list. All taxa in our sources that were not identified to species level or below, were removed from the comparison with the national backbone, as the national backbone contains only species-level and lower taxonomic ranks. We cross-referenced species on our list with the SANPC list to identify species classified as native on the SANPC to detect where extralimital alien species were included on lists. The number of extralimital species reported was compared between parks using a Fisher's exact test. We also compared the number of species listed per park to the number previously published for parks with a t-test and the contribution of different sources to current versus previous lists with a chi-square test. All statistical tests were conducted in R.

Results

Combined alien species list for all 21 South African National Parks

We recorded almost double the number of alien taxa within national parks in the current list compilation (1,330) compared to the previous compilation (752; Foxcroft et al. 2017), with a significant increase across parks ($t = 2.131$, $df = 18$, p-value = 0.047; Table 1). A further 170 species were identified from the catchment areas surrounding parks in research-grade iNaturalist data. Of the new species added to the overall list across parks, several publications on alien taxa, notably one on ornamental species in Kruger (Keet et al. 2022), contributed the highest

Table 1. The number of alien plant taxa documented in each national park and the number of species within different NEMBA legislative categories: Category 1 invasive species are considered to pose the largest threat (1a are prohibited and must be eradicated and 1b require compulsory control), Category 2 species are controlled by area and subject to permitting conditions and Category 3 species are subject to certain exemptions. Extralimital species, i.e. species that are native to parts of South Africa, but have been introduced to new areas, are reported separately. The number of plant taxa recorded in 2017 (Foxcroft et al. 2017) is provided for comparison.

Park	1a	1b	2	3	None	Extralimital	Total taxa	Total taxa (Foxcroft et al. 2017)
Addo Elephant	3	55	6	5	97	11	166	124
Agulhas		36	8	6	52	2	102	82
Augrabies Falls		14	2	1	28	2	45	54
Bontebok		42	6	1	79	24	128	80
Camdeboo	2	35	4	2	16	1	59	45
Garden Route	3	91	12	15	210	20	331	251
Golden Gate Highlands		36	9	7	46	2	98	77
Grasslands		9	4		12	0	25	NA
Kalahari Gemsbok		7			7	2	14	18
Karoo	1	17	2		9	0	29	22
Kruger	2	110	7	15	567	69	701	363
Mapungubwe		31	2	1	21	1	55	40
Marakele		19	3	2	4	0	28	20
Meerkat		1			3	0	4	NA
Mokala		12	1		6	0	19	27
Mountain Zebra		28	3		53	9	84	101
Namaqua		14	1		16	0	31	18
Richtersveld		9	1	1	13	1	24	15
Table Mountain	9	117	16	26	398	61	566	243
Tankwa Karoo		16	3		20	0	39	27
West Coast	1	16	3	3	22	1	45	15
TOTAL UNIQUE TAXA	**14**	**202**	**21**	**42**	**1066**	**206**	**1330**	**752**

number of new species to the list across parks (323). iNaturalist data added a further 124 species to the list, contributing an additional 9.3% to the total listed taxa. The scope of work recorded in the WIMS database has also increased since the previous publication, with 274 unique taxa represented in this source, 111 of which were new to this list and not included in other sources. For some parks, the number of taxa documented decreased in comparison to previous lists. At a park level, publications, largely those dedicated to listing alien species, contributed the highest number of taxa to lists. WIMS data and iNaturalist data both contributed significant amounts of new data to park-level lists (Fig. 4).

Of the species on the list, 271 (20%) were listed in the NEMBA regulations, with the majority of these in category 1b, requiring mandatory control. Thirty-six of the listed species had different legislative requirements in different parks. The average proportion of extralimital species included on park lists was 5% and seven park lists had no extralimital species, with a significant difference between the number of extralimital species included on lists for different parks (Fisher's exact test using 10,000 simulation replicates, $p < 0.0001$). Table Mountain National Park was of particular interest due to the potential impact of extralimital species on endemic species and we detected significantly more extralimital species on the list for Table Mountain compared to other parks (X-squared = 7.4761, df = 1, p-value = 0.006).

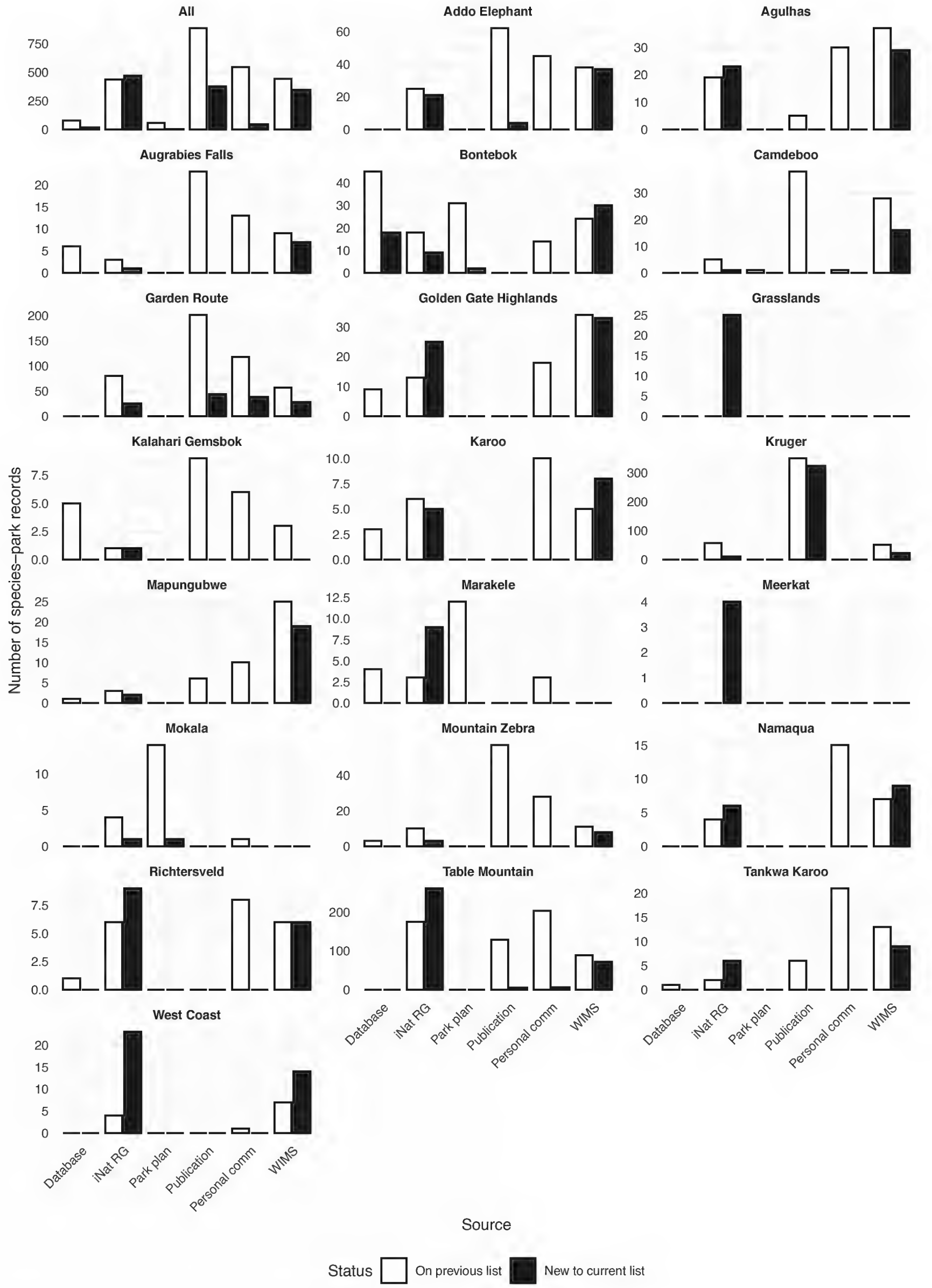

Figure 4. The number of unique species–park records contributed by each source in the current park species lists, categorised by whether the species-park combination was present in the previous list ("On previous list") or newly added during the current process ("New to current list"), across all parks and for each individual park. Note that y-axes vary across sub-plots. Across parks, the contribution of new records to parks differed significantly by source (X-squared = 390.34, df = 5, p-value < 2.2e-16). Sources: Database (unpublished data curated at park level or similar), iNat RG (iNaturalist Research Grade data), Park plan (species lists from Park Management Plans), Publication (species lists from literature), Personal comms (species included on lists by local experts), WIMS (Working for Water Information Management System data).

In total, across parks and their adjacent catchments, we recorded 1,500 unique taxa. Of these, 1,419 were species (or subspecies or lower taxonomic rank). We matched these species to the national backbone (SANBI 2025) using GBIF taxonKeys and 447 (31.5%) were not found. The proportion of species missing from the backbone did not differ between the park list (31.9%) and the watchlist of species recorded only in the adjacent catchments (28.8%, X2 = 0.50846, df = 1, p-value = 0.4758). Suppl. material 5: Checklist of species contains the complete list of alien species present in parks, the complete list of species removed and the complete species watchlist.

Challenges encountered during list compilation

The process we implemented was somewhat iterative in that, as we introduced certain solutions, new problems were encountered. Here, we present an overview of the challenges encountered and provide more detail of their resolution in the discussion.

Difficulty matching names across sources

We encountered several difficulties in matching names across sources due to spelling errors and the use of outdated names (synonyms). In formal data sources (e.g. publications, taxonomic databases), there were not many instances of spelling errors. However, management data had numerous spelling errors. In some instances, the automated name-cleaning process resolved these errors; however, 281 records, representing 54 taxa required manual matching or cleaning. There were two instances in which fuzzy matching resulted in an incorrect name. Management data had listed the common name Banana, which GBIF matched to *Banara*, a genus that does not occur locally in that protected area (Garden Route National Park). In the second case, for the same protected area, a spelling error "*Cinnamomum camphoratus*", most likely referring to *Cinnamomum camphora* (L.) J.Presl (present in the park), was matched to *Cinnamomum camphoratum* Blume, which is a synonym for *Cinnamomum sintoc* Blume, which does not occur in the protected area. Approximately 11% (157) of the names used across sources were initially identified to be synonyms on GBIF. Similarly, of 380 NEMBA-listed plant species, 41 (10.8%) names were deemed synonyms by GBIF.

Incorrect use of names

At the park level, 293 listings representing 236 unique taxa (species or genera) were deemed by experts to have been misidentified or incompletely identified (e.g. only to genus where the species was known) in a particular park and were marked for exclusion on the final list. Overall, 77 taxa were removed from the final list. Our data sources included records where species had been sent to herbaria for identification. Some taxa (e.g. *Eucalyptus* and Cyperaceae) could often not be identified by the herbaria. These records were, where possible, identified by comparing specimens collected in the field with species planted by, for example, plantation forestry historically, but several listings remain grouped by genus (e.g. *Eucalyptus* sp.).

A special case of misidentification is misapplied names. While plants may be correctly identified as alien and according to the names used in South Africa, in families such as the Cactaceae, where the taxonomy can be complicated, the names

of some species have been misapplied (Paterson et al. 2021). For example, what has been referred to as *Opuntia humifusa* (Raf.) Raf. in South Africa, is *Opuntia cespitosa* Raf., as the latter species was introduced to South Africa, but has been referred to by the incorrect name locally. *Cyrtomium falcatum* (L.fil.) C.Presl was misapplied in the Garden Route as *Cyrtomium micropterum* (Kunze) Ching, but a recent dissertation on ferns revealed this to be an error (Jones 2019).

Change in residence status and other errors

Some species were noted to change from alien to native (e.g. *Phytolacca icosandra L.* and some *Solanum* sp.). These cases were checked against the SANPC. While the latter backbone does not contain all the alien species, when it indicates a species as native, it was accepted as native. We did not thoroughly assess spatial accuracy. We recognise that errors in geographic coordinates associated with occurrence records can introduce errors into species lists and have, therefore, included spatial accuracy as a key component of the envisaged future database structure (see Suppl. material 1). In addition, we encountered a small number of cases where managers suggest that species have been removed from a park and, as such, should not be listed.

Discussion

We present a 10-step method for integrating species occurrence data from multiple sources to enable the generation of a dynamic database from which living species checklists can be drawn. Using data from South Africa's national parks, we trialled this approach and produced updatable checklists for all 21 SANParks-managed parks. Our park-wide checklist is significantly longer than previous versions due to the inclusion of several new comprehensive park-level studies, as well as iNaturalist data, although changes in the inclusion of species across invasion status (e.g. inclusion of ornamental or cultivated species, such as those listed in Keet et al. (2022)) may also have contributed to the increase. When evaluating individual park lists, however, our lists were shorter for some parks, reflecting the removal of misidentified species or synonyms from earlier lists, as well as our potentially incomplete use of available data. Changes in reported alien species numbers can be misleading (McGeoch et al. 2012; Rodrigues et al. 2022) and provide challenges for long-term monitoring or reporting, such as for the triennial national status reports (see Zengeya et al. 2026). An increase might suggest control failure or new invasions, when in fact, it may reflect expanded data sources or broader inclusion criteria. A decrease could appear to signal successful eradication, though it may result from implementing stricter listing criteria, removal of synonyms or other taxonomic duplication or source exclusions.

As indicators are increasingly used to track conservation outcomes (CBD 2022; McGeoch et al. 2023), it is crucial to address such inconsistencies. We recommend that indicators include guidance or interpretation on: (1) how alien species statuses and invasion stages were included (Richardson et al. 2000), especially in the treatment of cultivated species; (2) the inclusion/exclusion of extralimital species; and (3) verification processes, especially for non-expert and spatially imprecise data (Rodrigues et al. 2022). For this paper, we included all species known to be introduced from another country or another part of the country, but not native bush encroacher species. Such species would need to be included in the defined scope and categorised appropriately within the data collection framework.

We also developed water catchment-based species lists to provide information for park-specific watchlists. However, these should be interpreted alongside pathway data (Wilcox et al. 2024), as not all species disperse via waterways and the use of buffer zones around parks may need adjustment, based on pathway relevance. For example, Table Mountain National Park's watchlist included 137 additional species from surrounding catchments, but many of these are not linked to waterborne pathways. Less stringent inclusion criteria (e.g. use of non-research-grade iNaturalist data) could also increase the number of species of concern in the buffer zone. Non-research-grade data were excluded here due to frequent occurrence of misidentification in unverified data (Wenk et al. 2024). However, data are continually being updated, so records that may not have been research grade at the time of our download have since been reclassified. For example, iNaturalist research-grade records in our original drawdown (data uploaded between 15 July 2023 and 18 March 2024) increased from 25,170 to 29,716 (18%) when re-downloaded a year later (April 2025).

A key challenge in the process of species list compilation is that the overall workflow can never be fully automated (Reyserhove et al. 2020). Manual checks were required and will remain a requirement at several steps to identify cases of misidentification. For example, special characters, particularly associated with author names and names for hybrid species (e.g. the superscript in *Populus* × *canescens* (Aiton) Sm. or the Spanish tilde in *Hypochaeris brasiliensis* Larrañaga), resulted in no match to an accepted name, where an accepted name does exist. While the automation procedures we trialled reduced the occurrence of unmatched names, manual checking was still required. Any manual verification is a time-consuming process, with the potential for introducing human error (McGeoch et al. 2012). Given that integration of disparate historic datasets with newer data is complicated, it requires dedicated and substantial investment in time and resources, which is one of the main reasons that historic lists and datasets are overlooked.

Improving listing and reporting on alien species

Scope (inclusion of cultivated, planted or horticultural species)

Step 1 (Fig. 2) is crucial to a successful listing process. Clearly defining the geographic and taxonomic scope improves both the accuracy and usefulness of the list (Wilcox et al. 2024). A common challenge is the inclusion of planted species and effort is required to define and filter out unwanted records (Pyšek et al. 2012; Rebelo et al. 2025). Here, we included all existing alien species lists for parks to ensure these data exist in our database, but for management reasons, many of these happen to be ornamental species. While many planted individuals pose no risk of invasion, some do, especially when planted in large numbers (e.g. plantation forestry). For example, in parks like Garden Route, alien plant species were found in forest villages, picnic sites and pastures where they had naturalised or spread from refuse dumps. Cultivated plants that never escaped (e.g. carrots or potted begonias) have, however, not been included on lists. In Kruger National Park, data collected from staff and tourist areas included cultivated and horticultural species that were clearly marked as such (Keet et al. 2022). By annotating records with invasion status, from planted to invasive or transformer (Richardson et al. 2000), users of the database will be able to tailor lists to specific purposes. Including information on invasion extent

further supports accurate status assignment (see Pyšek et al. 2012 as an example of how species distribution was considered in listing in the Czech Republic). In parks like Table Mountain, where local biodiversity and endemism are high, extralimital species may pose a significant threat of hybridisation (Foxcroft et al. 2019). In such instances, including extralimital species on the list of invasions may be important for management. We note that, while there were significantly more extralimital species on the list for Table Mountain, many other parks have not included this scope. Depending on the purpose of the list, the geographic scope may be broader or narrower. For example, in some sources, species that clearly had the ability to spread (move through the landscape) were added to the species list when detected within 500 m from the park boundary.

Sources and detecting new species arrivals

One of the most important uses of alien species lists is to detect new invasions and potential invasion pathways (Essl et al. 2015; Faulkner et al. 2020). Doing this requires a thorough understanding of historical listing information as well as sources of new information. For example, citizen science and platforms like iNaturalist are becoming increasingly useful for the preparation of species lists and early detection of invasions (Richardson and Potgieter 2024; Gildenhuys et al. 2025). There are, however, some caveats. Even by using only research-grade plant identifications from iNaturalist, we still encountered numerous misidentifications. Classifying alien species from iNaturalist into a system that conceptualises the 'naturalisation to invasion' process (Richardson et al. 2000) is required to determine which species are 'Casual alien plants' and those that pose a larger threat like 'Naturalised, Invaders and Transformers'. While the latter classes may be of concern, 'Casual Aliens' may be cultivated and of little concern. Unless an iNaturalist observation is inside a park and for a species with known ability to spread, adding it to a species list would not be advisable as it may become necessary to report on its management (van Wilgen et al. 2025) in future and mislead data users and decision-makers. We suggest that any legislatively listed and/or 'Casual Alien' species close to, but outside the PA, go on to a 'watchlist' and be closely monitored for signs of spread.

The ability of citizen science and research projects to contribute to baseline occurrence data can be improved through the use of special data collection platforms, such as iNaturalist projects (Theobald et al. 2015). These can be set up to collect the data and metadata required for maintaining a dynamic database (Fig. 2, Steps 3–5) (Spear et al. 2023). Researchers working in protected areas could be required to upload species occurrence data to a project as a component of their research permit to enable integration of new data. Inclusion of citizen-science data comes with many other caveats, pros and cons (Larson et al. 2020; Wenk et al. 2024), but discussing these in detail is beyond the scope of this paper. Generally, while the identifications on iNaturalist may not be accurate, the photographs that accompany records can still be useful, especially in flagging observations where field verification is merited. An especially important field to take note of is the one that indicates if a record represents a planted individual. Depending on the purpose of the list, such records might be filtered out before downloading data (Rebelo et al. 2025).

Other than actual changes in species presence (true change in invasion), factors such as re-interpretation of data and updates in confidence in particular data can also affect the number of species listed in an area (Zengeya et al. 2026).

Our proposed use of columns indicating confidence in observations will allow these changes to be tracked in parallel with true invasion changes to provide information for reporting of likely new arrivals.

Matching names across sources

Matching names across sources can be challenging for a variety of reasons and, depending on the information available, may not always be resolvable. Firstly, there are spelling errors (Fig. 1, part 1a). Spelling errors and unusual characters can be managed by keeping both a Verbatim name and a cleaned name (= Taxon name) column in the occurrence database, preserving the original, while enabling standardisation. While playing a key component in automation, fuzzy matching also introduced errors, especially when it matched a misspelled name to the synonym of an unrelated species. Cases of fuzzy matching must be checked manually to avoid erroneous list additions (see results for examples). We recommend using scientific names rather than common names wherever possible. Secondly, taxonomic changes result in the use of synonyms that vary widely in their use between sources (Fig. 1, part 1b). The use of outdated names in management plans, legislation and national reporting tools adds further complication.

Incorrect use of names

Incorrect use of names may stem from misidentification or where names have been misapplied (Fig. 1, part 2). For misidentifications, including an identification accuracy column with annotations (e.g. herbarium verified, expert verified, photo ID, unverified, misidentification, misapplied name) allows filtering of uncertain records. A unique challenge that arose in management data was the use of surrogate names for species with similar clearing requirements. This occurred because WIMS, the system used for calculating work specifications, based on species traits (Neethling and Shuttleworth 2013), did not include clearing norms for all species. For instance, *Acacia baileyana* was used in place of *Banksia speciosa* in a clearing work contract. We updated these entries, based on notes on the contracts, such that the contract Verbatim name is retained, but the Taxon name is updated to the correct name with appropriate annotations in the comments column. Matching between the databases to enable future use of these name corrections will be possible, based on unique record identifiers. We also encountered misapplied names, such as the use of *Opuntia humifusa* in Kruger National Park. It was initially misidentified as *O. stricta*, later confirmed to be *O. cespitosa* (Foxcroft et al. 2023). This was detected as a name misapplication and the name later relisted as *O. humifusa*. This has, however, resulted in inconsistency, as management plans still refer to the incorrect name. Unlike the case of synonyms, misapplied names and misidentifications cannot be resolved via unique taxon identifiers and must be corrected manually (Rodrigues et al. 2022). All misapplied names should be annotated in the occurrence database and, where appropriate, in the source data to prevent recurring errors.

Name resolution remains difficult for some taxa (Rodrigues et al. 2022). While literature searches and local iNaturalist records may help, some genera (e.g. *Eucalyptus*) or families (e.g. Myrtaceae, Cyperaceae) require detailed review of floral traits for identification (Blakely 1955). In such cases, assigning genus-only labels

may be more accurate. Species-level assignment amidst uncertainty should only be attempted when necessary for data consolidation (e.g. reporting trends over time where a species is named at one time and a genus at an earlier time) and should be guided by user judgement.

Other listing applications and considerations

Spatial error

Geographic errors in data may arise for a variety of reasons, for example, GPS error or human error in interpretation of place names. These errors may result in a species inadvertently appearing inside or outside of a protected area, where the opposite is true (Wenk et al. 2024). Where source data are being cleaned (Fig. 2, Step 2), annotation of the accuracy of spatial information is useful for future filtering out of such errors. The current BIMS captures spatial precision, but while GPS coordinates may be precise, coordinates alone do not provide an indication of the accuracy associated with the observation. The BIMS is currently being revised to allow for more detailed capture relating to the verification of spatial accuracy in relation to observations. We propose the following options for discriminating spatial error in conjunction with precision of localities: GPS point, nBal Centroid (SANParks specific), Park section Centroid, Park Centroid, QDS Centroid, Inaccurate GPS location, Verified place name, Historical range map, Unverified place name or Unknown (see Suppl. material 1).

Extirpation or eradication of species

Extirpation or eradication (in ideal situations) is the management aim for many alien species in protected areas (Sankaran et al. 2023). Though it is rarely achieved, a process is needed to annotate records where individuals, populations or species have been successfully removed from a park. However, some level of caution is required in declaring a species eradicated, similar to declaring a species extinct on the IUCN Red List (Matthys 2023). In large parks, species may be cleared in one area, but re-appear elsewhere, possibly from new invasions or from seedbanks formed by the initial invasions. While some species, especially those present in small numbers or targeted by effective clearing methods (Reaser et al. 2020), may be extirpated or even eradicated, they remain on lists due to the absence of clear protocols for removing species from lists. Documenting true extirpations and/or eradications is important for evaluating management success. Including the date of last record in checklists is useful for users to assess current presence more accurately, but annotating eradication or extirpation in relation to species records requires additional guidance.

In some instances, species should be removed from lists because they were on the lists erroneously to begin with. Matthys et al. (in press) provide a protocol for evaluating evidence that would allow managers to remove species from lists. For example, in our revision of WIMS data, many species were noted to have been misidentified and updates suggested by experts (e.g. in the Garden Route, the species listed as *Datura ferox* L. should be *D. stramonium* L., while other species such as *Alnus viridis* (Chaix) DC. and *Senegalia (Acacia) caffra* (Thunb.) P.J.H.Hurter & Mabb. were noted to be unlikely and scored as Low confidence for the park). While these updates may affect the number of listed species, annotating the re-

cords as incorrect identifications or localities in the occurrence database and marking them for exclusion (at user discretion) for when lists are generated allows those tracking changes over time (Zengeya et al. 2026) to correctly interpret the changes.

Risk assessments, pathways of entry, interface with adjacent land-use

In addition to compiling species lists, risk assessments play a crucial role in checklist development. These assessments evaluate factors such as species-pathway associations and environmental suitability to predict possible invasion dynamics that can be useful for providing information for scenarios under climate change (Sankaran et al. 2023). By integrating this information into the checklist framework, conservation managers can prioritise management actions, based on the potential impact of alien species on native ecosystems. In protected areas, these strategies must also include a thorough understanding of the pathways through which alien species are introduced and spread. This involves identifying both biotic and abiotic pathways that facilitate invasions, such as human activity, trade, transport vessels and natural dispersal mechanisms (Foxcroft et al. 2019). In the current list compilation, we have generated a watchlist, based on plant species in catchments surrounding parks, but more nuanced buffers should also be considered in relation to other pathways. Watchlists could, for example, consider which species are being sold in local nurseries or pet stores or the likely impact of climate change (Lieurance et al. 2025) on existing emerging and invasive species (Faulkner et al. 2020; van Wilgen et al. 2023). The nature of land-use in areas surrounding a park (i.e. the extent of transformation, fragmentation and common practices) will assist in providing information for the most likely pathways of entry to a national park (Spear et al. 2013).

Interoperability and other listing processes

GBIF contains one of the largest compilations of species occurrence data globally with over 3 billion records (as of October 2025). Reyserhove et al. (2020) describe a protocol for publishing source data to GBIF via Darwin Core Archives, enabling interoperability and broad scientific reuse of biodiversity data. This application is ideal for researchers who wish to make their data available to the broader scientific community. However, protected area managers frequently have very specific data needs and may require nuances in data that are not provided for in general platforms like GBIF. This may include factors like local invasion status, local management priority and cross-references to other locally used management reference columns, such as site names, site codes and management codes. Locally curated databases also enable users to record both international (e.g. GBIF) and nationally used unique taxon identifiers to enable matching across systems. Our process provides a detailed workflow for developing and maintaining occurrence databases for local use, structured to operate in a manner comparable to broader open-data platforms like GBIF. Such internal management systems provide the agility to alter database structure as needs arise, enabling locally relevant management functionality, while remaining interoperable with global systems (Pagad et al. 2015). To ensure interoperability, local protected area databases should retain an internationally recognised unique taxon identifier, such as the GBIF taxonKey incorporated in our system, along with the Darwin Core fields required by platforms like GBIF (see Suppl. material 1). In the context of alien species, the Global Register of

Introduced and Invasive Species (GRIIS, Pagad et al. 2018) provides a platform for maintaining national alien species checklists. GRIIS also uses the GBIF taxonomic backbone and provides a mechanism for ensuring consistency in alien species reporting between local, national and international scales. By adopting the GBIF taxonKey and Darwin Core compatible fields, any species list generated using our proposed process will be interoperable with GBIF and GRIIS.

Conclusions

Developing a user-friendly, relational database with standardised fields, regular updates and institutional support is critical to enabling conservation staff to manage biodiversity data effectively (Michener et al. 2011). Standardisation using a taxonomic backbone, such as GBIF, can streamline data sharing and enhance conservation decision-making, particularly where the standardisation occurs across national and international data systems (Pagad et al. 2018; Thomson et al. 2021; Spear et al. 2023). To achieve what we define as a *living checklist*, six essential criteria must be met (see also Reyserhove et al. 2020), including: (1) use of original data sources over derived ones when setting up a database; (2) assignment of globally unique record identifiers to each occurrence record to enable traceability across source and derived datasets (see Suppl. material 1: Database collection template); (3) retention of both Verbatim and cleaned Taxon names to facilitate accurate matching; (4) annotation of records with identification and spatial accuracy; (5) selection of an appropriate taxonomic backbone, based on user context, paired with the ability to record both the backbone version of the Taxon name and its associated taxonKey and the accepted name currently associated with this name if it is a synonym. In addition to this, our process provides a taxonKey for the accepted name, which we used in aggregation of species lists; and finally, (6) documentation of the invasion status of each taxon which allows for informed decisions on inclusion of taxa on lists. Depending on the use of the list, the required database structure and backbones may differ slightly. Here, we chose to use GBIF as it is used in the SANParks BIMS, as well as the GRIIS, but other backbones may have closer matches to nationally understood taxonomy. Meeting the above criteria allows integration of static historical lists with dynamic platforms, such as GBIF and iNaturalist and establishes a framework for incorporating future records. Citizen-science platforms and curated iNaturalist projects offer scalable, cost-effective solutions for ongoing data collection. Through coordinated efforts across researchers, policy-makers and protected area managers, living checklists can underpin the indicators used to address the challenges of invasive species and biodiversity loss in protected areas.

Acknowledgements

We thank Judith Botha, Nazley Liddle, Kylen Brown and Murunwa Mathoho for guidance on the SANParks BIMS and, along with Anna Brooke, for assisting with the cleaning of some data sources and trialling the initial name-matching process. We thank ChatGPT and its creators (OpenAI 2025) for assistance in the refinement of functions to allow for faster processing of large datasets. Lesley Hendersen is thanked for commenting on inaccuracies in the species lists. Comments from the editors and three anonymous reviewers significantly improved the manuscript. Dian Spear was funded by JRS Biodiversity Foundation (grant 60916).

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Use of AI

AI was used in the refinement of functions to allow for faster processing of large datasets.

Funding

The development of the BIMS was funded by the JRS Biodiversity Foundation. Updates to the BIMS identified during this listing process are being funded by the Agence Française de Développement (grant CZA 1171 01 V).

Author contributions

Conceptualization: DS, NC, NJJW, JB. Data curation: EJW, CC, LCF, NJJW, NC. Formal analysis: NJJW, EJW. Funding acquisition: DS. Methodology: JB, DS, NJJW. Project administration: NJJW. Supervision: LCF, JB. Validation: DS, JB. Visualization: EJW. Writing - original draft: NJJW, DS, EJW. Writing - review and editing: NJJW, JB, LCF, NC, EJW, CC.

Author ORCIDs

Nicola J. van Wilgen https://orcid.org/0000-0001-8110-698X
Emma J. Wright https://orcid.org/0009-0005-3853-9225
Johan Baard https://orcid.org/0000-0002-8871-4184
Bevin Ahrends https://orcid.org/0009-0002-7911-5267
Nicholas Cole https://orcid.org/0009-0007-2960-5221
Chad Cheney https://orcid.org/0000-0003-4536-2163
Llewellyn C. Foxcroft https://orcid.org/0000-0002-7071-6739
Dian Spear https://orcid.org/0000-0002-2417-3980

Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

References

Ackland SJ, Richardson DM, Robinson TB (2024) A method for conveying confidence in iNaturalist observations: A case study using non-native marine species. Ecology and Evolution 14: e70376. https://doi.org/10.1002/ece3.70376

Anderson EF (2001) The cactus family. Timber Press, Portland.

Baard JA, Kraaij T (2014) Alien flora of the Garden Route National Park, South Africa. South African Journal of Botany 94: 51–63. https://doi.org/10.1016/j.sajb.2014.05.010

Baard JA, Kraaij T (2019) Use of a rapid roadside survey to detect potentially invasive plant species along the Garden Route, South Africa. Koedoe 61: a1515. https://doi.org/10.4102/koedoe.v61i1.1515

Blakely WF (1955) A key to the Eucalypts. 2nd edn. Foresty and Timber Bureau, Canberra.

Bromilow C (2010) Problem plants and alien weeds of South Africa. Briza Publications CC, Pretoria.

Bromilow C (2014) Problem plants and alien weeds of southern Africa, revised fourth edition. Briza Publications, Pretoria, South Africa.

Carrero C, Beckman Bruns E, Frances A, Jerome D, Knapp W, Meyer A, Mims R, Pivorunas D, Speed D, Treher Eberly A, Westwood M (2022) Data sharing for conservation: A standardized checklist of US native tree species and threat assessments to prioritize and coordinate action. Plants, People, Planet 5: 600–616. https://doi.org/10.1002/ppp3.10305

CBD (2022) Decision adopted by the conference of the parties to the convention on biological diversity 15/4. Kunming-Montreal global biodiversity framework. https://www.cbd.int/doc/decisions/cop-15/cop-15-dec-04-en.pdf

Chamberlain S, Barve V, Mcglinn D, Oldoni D, Desmet P, Geffert L, Ram K (2024) rgbif: Interface to the Global Biodiversity Information Facility API. R package version 3.8.1. https://CRAN.R-project.org/package=rgbif

Cheney C, Esler KJ, Foxcroft LC, van Wilgen NJ, McGeoch MA (2018) The impact of data precision on the effectiveness of alien plant control programmes: A case study from a protected area. Biological Invasions 20: 3227–3243. https://doi.org/10.1007/s10530-018-1770-8

Cheney C, Esler KJ, Foxcroft LC, van Wilgen NJ (2020a) The dominating influence of efficacy above management strategy in the long-term success of alien plant clearing programmes. Journal of Environmental Management 271: 110836. https://doi.org/10.1016/j.jenvman.2020.110836

Cheney C, Smith RJ, Becker R, Gerber L, Tyolo K, Esler K, Foxcroft L, van Wilgen N, McGeoch M (2020b) Quantifying range structure to inform management in invaded landscapes - Dataset. Journal of Applied Ecology 58(2): 338–349. https://doi.org/10.1111/1365-2664.13765

Cheney C, Wilgen NJ, Esler KJ, Foxcroft LC, McGeoch MA (2021) Quantifying range structure to inform management in invaded landscapes. Journal of Applied Ecology 58: 338–349. https://doi.org/10.1111/1365-2664.13765

Danihelka J, Chrtek JJ, Kaplan Z (2012) Checklist of vascular plants of the Czech Republic. Preslia 84: 647–811.

Droege S, Cyr A, Larivée J (1998) Checklists: An Under-Used Tool for the Inventory and Monitoring of Plants and Animals. Conservation Biology : The Journal of the Society for Conservation Biology 12: 1134–1138. https://doi.org/10.1046/j.1523-1739.1998.96402.x

Essl F, Bacher S, Blackburn TM, Booy O, Brundu G, Brunel S, Cardoso A-C, Eschen R, Gallardo B, Galil B, García-Berthou E, Genovesi P, Groom Q, Harrower C, Hulme PE, Katsanevakis S, Kenis M, Kühn I, Kumschick S, Martinou AF, Nentwig W, O'Flynn C, Pagad S, Pergl J, Pyšek P, Rabitsch W, Richardson DM, Roques A, Roy HE, Scalera R, Schindler S, Seebens H, Vanderhoeven S, Vilà M, Wilson JRU, Zenetos A, Jeschke JM (2015) Crossing Frontiers in Tackling Pathways of Biological Invasions. Bioscience 65: 769–782. https://doi.org/10.1093/biosci/biv082

Evans AE, Brewington L, Brown-Lima C, Fusco E, Gregg RM, Lieurance D, Parsons EW, Nagy RC, Thurman L, Morelli TL (2025) Challenges and priorities for climate-informed invasive species management across multiple scales. Conservation Science and Practice 7. https://doi.org/10.1111/csp2.70074

Faulkner KT, Burness A, Byrne M, Kumschick S, Peters K, Robertson MP, Saccaggi DL, Weyl OLF (2020) South Africa's pathways of introduction and dispersal and how they have changed over time. In: van Wilgen BW, Measey GJ, Richardson DM, Wilson JR, Zengeya TA (Eds) Springer, Cham, Switzerland, 311–352. https://doi.org/10.1007/978-3-030-32394-3_12

Foxcroft LC, Richardson DM (2003) Managing alien plant invasions in the Kruger National Park, South Africa. In: Child LE, Brock JH, Brundu G, Prach K, Pyšek P, Wade PM, Williamson M (Eds) Plant Invasions: Ecological threats and management solutions. Backhuys Publishers, Leiden, The Netherlands, 385–403.

Foxcroft LC, Henderson L, Nichols GR, Martin BW (2003) A revised list of alien plants for the Kruger National Park. Koedoe 46: 21–44. https://doi.org/10.4102/koedoe.v46i2.54

Foxcroft LC, Richardson DM, Wilson JRU (2008) Ornamental plants as invasive aliens: Problems and solutions in Kruger National Park, South Africa. Environmental Management 41: 32–51. https://doi.org/10.1007/s00267-007-9027-9

Foxcroft L, van Wilgen NJ, Baard J, Cole N (2017) Biological invasions in South African National Parks. Bothalia 47: a2158. https://doi.org/10.4102/abc.v47i2.2158

Foxcroft LC, Spear D, van Wilgen NJ, McGeoch MA (2019) Assessing the association between pathways of alien plant invaders and their impacts in protected areas. NeoBiota 43: 1–25. https://doi.org/10.3897/neobiota.43.29644

Foxcroft LC, Moodley D, Nichols GR, Pyšek P (2023) Naturalized and invasive alien plants in the Kruger National Park, South Africa. Biological Invasions 25: 3049–3064. https://doi.org/10.1007/s10530-023-03098-0

Gaiji S, Chavan V, Ariño AH, Otegui J, Hobern D, Sood R, Robles E (2013) Content assessment of the primary biodiversity data published through GBIF network: Status, challenges and potentials. Biodiversity Informatics 8: 94–172. https://doi.org/10.17161/bi.v8i2.4124

Geldenhuys CJ (1993) Floristic composition of the southern Cape forests with an annotated check-list. South African Journal of Botany 59: 26–44. https://doi.org/10.1016/S0254-6299(16)30771-2

Gildenhuys CP, Potgieter LJ, Hui C, Richardson DM (2025) Deriving inventories of non-native plant species from iNaturalist: Insights from urban centres of the Western Cape, South Africa. ARPHA Preprints 6: e156050. https://doi.org/10.3897/neobiota.104.155832

Govaerts R (2025) World Checklist of Vascular Plants (WCVP) – v14 28th May 2025. Board of Trustees of the Royal Botanic Gardens, Kew, Kew, UK.

Grenié M, Berti E, Carvajal-Quintero J, Dädlow GML, Sagouis A, Winter M (2022) Harmonizing taxon names in biodiversity data: A review of tools, databases and best practices. Methods in Ecology and Evolution 14: 12–25. https://doi.org/10.1111/2041-210x.13802

Hamer M, Victor J, Smith GF (2012) Best practice guide for compiling, maintaining and disseminating national species checklists, version 1.0, released in October 2012. Global Biodiversity Information Facility, Copenhagen, 40 pp. http://www.gbif.org/orc/?doc_id=94752

Henderson L (1995) Plant invaders of South Africa. A pocket field guide to the identification of 161 of the most important and potentially important alien species. Handbook No. 5. Plant protection Research Institute, Agricultural Research Council, Pretoria.

Henderson L (2001) Alien Weeds and Invasive Plants: A Complete Guide to Declared Weeds and Invaders in South Africa. Plant Protection Research Institute Handbook no. 12. Agricultural Research Council, Pretoria, South Africa.

Henderson L (2020) Invasive alien plants in South Africa. ARC-PPRI Handbook No. 21.

Henderson L, Cilliers CJ (2002) Invasive aquatic plants: a guide to the identification of the most important and potentially dangerous invasive aquatic and wetland plants in South Africa. Plant Protection Research Institute Handbook No. 16, Agricultural Research Council, Pretoria, South Africa.

IPBES (2023) Summary for Policymakers of the Thematic Assessment Report on Invasive Alien Species and their Control. In: Roy HE, Pauchard A, Stoett P, Renard Truong T, Bacher S, Galil BS, Hulme PE, Ikeda T, Sankaran KV, McGeoch MA, Meyerson LA, Nuñez MA, Ordonez A, Rahlao SJ, Schwindt E, Seebens H, Sheppard AW, Vandvik V (Eds) Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES secretariat. [Bonn, Germany] https://doi.org/10.5281/zenodo.7430692

Jones EJ (2019) Terrestrial alien ferns (Polypodiophyta): A global assessment of traits associated with invasiveness and their distribution and status in South Africa. MSc thesis, Port Elizabeth: Nelson Mandela University.

Keet J-H, Datta A, Foxcroft LC, Kumschick S, Nichols GR, Richardson DM, Wilson JRU (2022) Assessing the level of compliance with alien plant regulations in a large African protected area. Biological Invasions 24: 3831–3844. https://doi.org/10.1007/s10530-022-02883-7

Klein H, Zimmermann HG (2020) Invasive cacti in South Africa: Their identification and control Agricultural Research Council.

Larson ER, Graham BM, Achury R, Coon JJ, Daniels MK, Gambrell DK, Jonasen KL, King GD, LaRacuente N, Perrin-Stowe TIN, Reed EM, Rice CJ, Ruzi SA, Thairu MW, Wilson JC, Suarez AV (2020) From eDNA to citizen science: Emerging tools for the early detection of invasive species. Frontiers in Ecology and the Environment 18: 194–202. https://doi.org/10.1002/fee.2162

Lieurance D, Canavan S, Faulkner KT, O'Shaughnessy KA, Lockwood JL, Parsons EW, Avery JD, Daniel W (2025) Understanding and managing introduction pathways into protected areas in a changing climate. Biological Invasions 27: 74. https://doi.org/10.1007/s10530-025-03534-3

Manes S, Costello MJ, Beckett H, Debnath A, Devenish-Nelson E, Grey K-A, Jenkins R, Khan TM, Kiessling W, Krause C, Maharaj SS, Midgley GF, Price J, Talukdar G, Vale MM (2021) Endemism increases species' climate change risk in areas of global biodiversity importance. Biological Conservation 257: 109070. https://doi.org/10.1016/j.biocon.2021.109070

Masubelele ML, Foxcroft LC, Milton SJ (2009) Alien plant species list and distribution for Camdeboo National Park, Eastern Cape Province, South Africa. Koedoe 51: a515. https://doi.org/10.4102/koedoe.v51i1.515

Matthys CK (2023) The development of a protocol for declaring alien species absent from South Africa. Cape Town: Cape Peninsula University of Technology.

Matthys C, Wilson JRU, Geerts S (in press) Argument maps can support decisions to declare the presence of alien species: South Africa as a case study. NeoBiota.

McGeoch MA, Spear D, Kleynhans EJ, Marais E (2012) Uncertainty in invasive alien species listing. Ecological Applications : A Publication of the Ecological Society of America 22: 959–971. https://doi.org/10.1890/11-1252.1

McGeoch MA, Ordonez A, Howard PL, Groom QJ, Shrestha BB, Fernandez M, Brugnoli E, Bwalya B, Byun C, Ksenofontov S, Ojaveer H, Simberloff D, Mungi NA, Rono B (2023) Chapter 6: Governance and policy options for the management of biological invasions. In: Roy HE, Pauchard A, Stoett P, Renard Truong T (Eds) Thematic Assessment Report on Invasive Alien Species and their Control of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES secretariat, Bonn, Germany. https://doi.org/10.5281/zenodo.7430747

Michener WK, Porter J, Servilla M, Vanderbilt K (2011) Long term ecological research and information management. Ecological Informatics 6: 13–24. https://doi.org/10.1016/j.ecoinf.2010.11.005

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403: 853–858. https://doi.org/10.1038/35002501

Natural Resources Conservation Service (2025) PLANTS Database. United States Department of Agriculture. [Accessed September 2025 at https://plants.usda.gov]

Neethling H, Shuttleworth B (2013) Revision of the working for water workload norms. Forestry Solutions, White River.

Nelufule T, Robertson MP, Wilson JRU, Faulkner KT (2022) Native-alien populations—An apparent oxymoron that requires specific conservation attention. NeoBiota 74: 57–74. https://doi.org/10.3897/neobiota.74.81671

OpenAI (2025) ChatGPT (March 2025 version). https://chat.openai.com/chat [accessed March 2025]

Pagad S, Genovesi P, Carnevali L, Scalera R, Clout M (2015) IUCN SSC Invasive Species Specialist Group: Invasive alien species information management supporting practitioners, policy makers and decision takers. Management of Biological Invasions: International Journal of Applied Research on Biological Invasions 6: 127–135. https://doi.org/10.3391/mbi.2015.6.2.03

Pagad S, Genovesi P, Carnevali L, Schigel D, McGeoch MA (2018) Introducing the Global Register of Introduced and Invasive Species. Scientific Data 5: 170202. https://doi.org/10.1038/sdata.2017.202

Paterson ID, Klein H, Muskett PC, Griffith TC, Mayonde S, Mofokeng K, Mnqeta Z, Venter N (2021) Biological Control of Cactaceae in South Africa. African Entomology 19(2): 230–246. https://doi.org/10.4001/003.029.0713

Pyšek P, Danihelka J, Sádlo J, Chrtek J, Chytry M, Jarošík V, Kaplan Z, Krahulec F, Moravcová L, Pergl J, Štajerová K, Tichý L (2012) Catalogue of alien plants of the Czech Republic (2nd edn): Checklist update, taxonomic diversity and invasion patterns. Preslia 84: 155–255.

Pyšek P, Hulme PE, Meyerson LA, Smith GF, Boatwright JS, Crouch NR, Figueiredo E, Foxcroft LC, Jarošík V, Richardson DM, Čuda J, Wilson JRU (2013) Hitting the right target: Taxonomic challenges for, and of, plant invasions. AoB Plants 5: plt042. https://doi.org/10.1093/aobpla/plt042

Reaser JK, Burgiel SW, Kirkey J, Brantley KA, Veatch SD, Burgos-Rodríguez J (2020) The early detection of and rapid response (EDRR) to invasive species: A conceptual framework and federal capacities assessment. Biological Invasions 22: 1–19. https://doi.org/10.1007/s10530-019-02156-w

Rebelo AG, Holmes PM, Dorse C, Wood J (2011) Impacts of urbanization in a biodiversity hotspot: Conservation challenges in Metropolitan Cape Town. South African Journal of Botany 77: 20–35. https://doi.org/10.1016/j.sajb.2010.04.006

Rebelo AG, Holmes PM, Spear D, Klopper RR, van Wilgen NJ (2025) Lessons learned from compiling a flora checklist for Table Mountain National Park. Koedoe 67: a1856. https://doi.org/10.4102/KOEDOE.v67i1.1856

Reyserhove L, Desmet P, Oldoni D, Adriaens T, Strubbe D, Davis AJS, Vanderhoeven S, Verloove F, Groom Q (2020) A checklist recipe: Making species data open and FAIR. Database : The Journal of Biological Databases and Curation 2020: baaa084. https://doi.org/10.1093/database/baaa084

Richardson DM, Potgieter LJ (2024) A living inventory of planted trees in South Africa derived from iNaturalist. South African Journal of Botany 173: 365–379. https://doi.org/10.1016/j.sajb.2024.08.006

Richardson DM, Pyšek P, Rejmánek M, Barbour MG, Panetta FD, West CJ (2000) Naturalization and invasion of alien plants: Concepts and definitions. Diversity & Distributions 6: 93–107. https://doi.org/10.1046/j.1472-4642.2000.00083.x

Rodrigues AV, Nakamura G, Staggemeier VG, Duarte L (2022) Species misidentification affects biodiversity metrics: Dealing with this issue using the new R package naturaList. Ecological Informatics 69: 101625. https://doi.org/10.1016/j.ecoinf.2022.101625

Roura-Pascual N, Richardson DM, Krug RM, Brown A, Chapman RA, Forsyth GG, Le Maitre DC, Robertson MP, Stafford L, Van Wilgen BW, Wannenburgh A, Wessels N (2009) Ecology and management of alien plant invasions in South African fynbos: Accommodating key complexities in objective decision making. Biological Conservation 142: 1595–1604. https://doi.org/10.1016/j.biocon.2009.02.029

Roux DJ, Kingsford RT, Cook CN, Carruthers J, Dickson K, Hockings M (2019) The case for embedding researchers in conservation agencies. Conservation Biology : The Journal of the Society for Conservation Biology 33: 1266–1274. https://doi.org/10.1111/cobi.13324

SANBI (2025) The South African National Plant Checklist: yearly release and official documentation. South African National Biodiversity Institute. https://doi.org/10.5281/zenodo.15050848 [Accessed April 2025]

Sankaran KV, Schwindt E, Sheppard AW, Foxcroft LC, Vanderhoeven S, Egawa C, Peacock L, Castillo ML, Zenni RD, Müllerová J, González-Martínez AI, Bukombe JK, Wanzala W, Mangwa DC (2023) Chapter 5: Management; challenges, opportunities and lessons learned. In: Roy HE, Pauchard A, Stoett P, Renard Truong T (Eds) Thematic Assessment Report on Invasive Alien Species and their Control of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES secretariat, Bonn, Germany. https://doi.org/10.5281/zenodo.7430747

SANParks (2024) Working for Water Information Management System data for the period 2003-2024. Unpublished database, Cape Town.

Singh N, Spear D, Baard J, Mapuru M, Bezuidenhout H, Van Wyk P, Brown K, Liddle N (2025) Pressed for impact: Unleashing the potential of South African National Parks plant records. Koedoe 67: a1868. https://doi.org/10.4102/koedoe.v67i1.1868

Smith G, Willis C (1999) Index herbariorum: southern African supplement. Southern African Botanical Diversity Network, Pretoria, 181 pp.

Spear D, Brown K, Liddle N, Parenzee CJ, van Wilgen NJ (2026) Optimising species checklists for protected areas in the digital age. Koedoe 68: a1858. https://doi.org/10.4102/koedoe.v68i1.1858

Spear D, Foxcroft LC, Bezuidenhout H, McGeoch MA (2013) Human population density explains alien species richness in protected areas. Biological Conservation 159: 137–147. https://doi.org/10.1016/j.biocon.2012.11.022

Spear D, McGeoch MA, Foxcroft LC, Bezuidenhout H (2011) Alien species in South Africa's national parks. Koedoe 53(1032): 1–1034. https://doi.org/10.4102/koedoe.v53i1.1032

Spear D, van Wilgen NJ, Rebelo AG, Botha JM (2023) Collating biodiversity occurrence data for conservation. Frontiers in Ecology and Evolution 11: 1037282. https://doi.org/10.3389/fevo.2023.1037282

Stephenson PJ, Bowles-Newark N, Regan E, Stanwell-Smith D, Diagana M, Hoft R, Abarchi H, Abrahamse T et al. (2017) Unblocking the flow of biodiversity data for decision-making in Africa. Biological Conservation 213: 335–340. https://doi.org/10.1016/j.biocon.2016.09.003

Stohlgren TJ, Pyšek P, Kartesz J, Nishino M, Pauchard A, Winter M, Pino J, Richardson DM, Wilson JRU, Murray BR, Phillips ML, Ming-yang L, Celesti-Grapow L, Font X (2011) Widespread plant species: Natives versus aliens in our changing world. Biological Invasions 13: 1931–1944. https://doi.org/10.1007/s10530-011-0024-9

Team RDC (2024) R: A language and environment for statistical computing. R Foundation for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Theobald EJ, Ettinger AK, Burgess HK, DeBey LB, Schmidt NR, Froehlich HE, Wagner C, HilleRisLambers J, Tewksbury J, Harsch MA, Parrish JK (2015) Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. Biological Conservation 181: 236–244. https://doi.org/10.1016/j.biocon.2014.10.021

Thomson SA, Thiele K, Conix S, Christidis L, Costello MJ, Hobern D, Nikolaeva S, Pyle RL, van Dijk PP, Weaver H, Zachos FE, Zhang Z-Q, Garnett ST (2021) Towards a global list of accepted species II. Consequences of inadequate taxonomic list governance. Organisms, Diversity & Evolution 21: 623–630. https://doi.org/10.1007/s13127-021-00518-8

Turbelin AJ, Malamud BD, Francis RA (2017) Mapping the global state of invasive alien species: Patterns of invasion and policy responses. Global Ecology and Biogeography 26: 78–92. https://doi.org/10.1111/geb.12517

van Wilgen BW, Cole NS, Baard J, Cheney C, Engelbrecht K, Stafford L, Turner AA, van Wilgen NJ, Wannenburgh AM (2025) Progress towards the control of invasive alien species in the Cape Floristic Region's protected areas. Biological Invasions 27: 8. https://doi.org/10.1007/s10530-024-03459-3

van Wilgen BW, Wannenburgh A, Wilson JRU (2022b) A review of two decades of government support for managing alien plant invasions in South Africa. Biological Conservation 274: 109741. https://doi.org/10.1016/j.biocon.2022.109741

van Wilgen BW, Zengeya TA, Richardson DM (2022a) A review of the impacts of biological invasions in South Africa. Biological Invasions 24: 27–50. https://doi.org/10.1007/s10530-021-02623-3

van Wilgen NJ, Faulkner KT, Robinson TB, South J, Beckett H, Janion-Scheepers C, Measey GJ, Midgley GF, Richardson DM (2023) Climate change and biological invasions in South Africa. In: Ziska LH (Ed.) Invasive Species and Global Climate Change, 2nd edn. CAB International, Wallingford UK, Boston USA, 158–187.

van Wilgen NJ, van Wilgen BW, Midgley GF (2020) Biological Invasions as a Component of South Africa's Global Change Research Effort. In: van Wilgen BW, Measey J, Richardson DM, Wilson JR, Zengeya TA (Eds) Biological Invasions in South Africa. Springer, Cham, Switzerland, 855–878. https://doi.org/10.1007/978-3-030-32394-3_29

Victor JE, Koekemoer M, Fish L, Smithies SJ, Mössmer M (2004) Herbarium essentials: The southern African herbarium user manual. SABONET, Pretoria.

Walters M, Figueiredo E, Crouch NR, Winter PJ, Smith G, Zimmermann HG, Mashope BK (2011) Naturalised and invasive succulents of southern Africa. Belgian Development Cooperation, Brussels, Belgium.

Watson JEM, Venegas-Li R, Grantham H, Dudley N, Stolton S, Rao M, Woodley S, Hockings M, Burkart K, Simmonds JS, Sonter LJ, Sreekar R, Possingham HP, Ward M (2023) Priorities for protected area expansion so nations can meet their Kunming-Montreal Global Biodiversity Framework commitments. Integrative Conservation 2: 140–155. https://doi.org/10.1002/inc3.24

Wenk E, Mesaglio T, Keith D, Cornwell W (2024) Curating protected area-level species lists in an era of diverse and dynamic data sources. Ecological Informatics 84: 102921. https://doi.org/10.1016/j.ecoinf.2024.102921

Wickham H, François R, Henry H, Müller K, Vaughan D (2023) dplyr: A Grammar of Data Manipulation. R package version 1.1.4. https://github.com/tidyverse/dplyr, https://dplyr.tidyverse.org

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7: e29715. https://doi.org/10.1371/journal.pone.0029715

Wilcox M, DiBacco C, Sardelis S, McKenzie C, Howland K, Therriault T (2024) A framework for the development of defensible and fit-for-purpose priority lists for non-indigenous species. Management of Biological Invasions: International Journal of Applied Research on Biological Invasions 15: 635–656. https://doi.org/10.3391/mbi.2024.15.4.11

Zengeya TA, Faulkner KT, Mtileni MP, Fernandez Winzer L, Kumschick S, McCulloch-Jones EJ, Miza-Tshangana SA, Robinson TB, Sifuba A, Engelbrecht W, van Wilgen BW, Wilson JRU (in press) A checklist of alien taxa for South Africa. NeoBiota.

Zengeya TA, Faulkner KT, Mtileni MP, Wilson JRU (2026) Lessons and challenges in creating alien species lists: insights from South Africa's national reports on the status and management of biological invasions. NeoBiota 101: 203-222. https://doi.org/10.3897/neobiota.101.162932

Supplementary material 1

Data collection template

Authors: Nicola van Wilgen, Dian Spear, Johan Baard, Nicholas Cole, Bevin Ahrends, Llewellyn Foxcroft, Emma Wright

Data type: xlsx

Explanation note: This is an excel spreadsheet that proposes the minimum fields required for data collection to support a dynamic occurrence database for use in checklists and indicators. The template indicates the fields required by GBIF and the names of comparable fields in GBIF where names differ.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.105.159101.suppl1

Supplementary material 2

R code for checking species names and matching data

Authors: Nicola van Wilgen

Data type: R file

Explanation note: This R code provides a mechanism to match large numbers of taxa names to GBIF to identify synonyms and provide taxonKeys. The code also indicates at which steps manual checking of outputs will be required.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.105.159101.suppl2

Supplementary material 3

NEMBA list categories for applying matching code

Authors: Nicola van Wilgen

Data type: zip

Explanation note: This file contains xlsx file (3a) and csv file (3b). Two files are provided within this supplementary material. The first (3a) represents the South African National NEMBA species regulation list of alien species and can be used in conjunction with the R code as an example of how to prepare outside databases for matching to species lists and occurrence data. The second (3b) is a CSV file that represents the cleaned output of this file that includes GBIF unique taxonKeys per species. The code in Suppl. material 2 can be used in conjunction with these files to generate park-specific lists.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.105.159101.suppl3

Supplementary material 4

Cleaned WIMS data and workflow

Authors: Nicola van Wilgen, Bevin Ahrends, Nicholas Cole

Data type: xlsx

Explanation note: This is an excel workbook comprised of data extracted from the Working for Water System (WIMS) for species cleared across all national parks in South Africa between 2004 and 2024. Name corrections have been applied as per the workflow described in the metadata. The workbook contains a record identification field that will enable matching between the original WIMS database and the dynamic occurrence database (BIMS).

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.105.159101.suppl4

Supplementary material 5

Checklist of species

Authors: Nicola van Wilgen, Johan Baard, Emma Wright, Bevin Ahrends, Nicholas Cole, Chad Cheney, Llewellyn C. Foxcroft, Dian Spear

Data type: xlsx

Explanation note: This is an excel workbook comprised of several sheets that include metadata as well as the output of applying the process described in this paper. The outputs include: (1) a list of all species documented per national park with details of sources and synonyms and (2) a list of all taxa that were removed from these lists per park based on expert opinion, with reasons for removal and a (3) watchlist of species that were not identified in parks, but documented in iNaturalist data from catchment areas around each park.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.105.159101.suppl5